Detection of base travel groups with different sensitivities to new high-speed rail services: Non-negative tensor decomposition approach

Detection of Base Travel Groups with Different Sensitivities to New High-speed Rail Services: Non-Negative Tensor Decomposition Approach

Hiromichi Yamaguchi[a,*], Shoichiro Nakayama[a]

[a]*Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa, Ishikawa, Japan*

Abstract

How many base travel groups (models) are necessary for clarifying the long-term day-to-day dynamics of intercity travel? In the past, several travel purposes (*e.g.*, sightseeing, business, etc.) have been assumed. However, mobile-phone location data enables us to answer the above question because of their detailed time-series information. In this study, we propose a method for deriving the basic travel groups necessary for clarifying the time-series changes by applying nonnegative tensor factorization (NTF). This method is applied to the time-series data of several long-distance travelers to the Ishikawa prefecture, to where the Hokuriku High-speed rail (HSR) has been newly extended. Based on this, the number of base travel groups necessary for predicting the effect of the new HSR is estimated as twelve, which is greater than the number used in the previous demand forecasting models. The estimated groups include components that appear to correspond to different travel purposes (e.g., sightseeing, business, and homecoming), as in previous surveys. These results confirm that the methodology proposed in this study can clearly extract groups with different elasticities, due to the traffic service. The HSR effect can be clarified by dividing it into several characteristics and detailed components. In addition, if multiple HSR effects are analyzed, a more accurate demand-forecasting model for the new HSR service can be proposed.

Keywords: High-speed rail, Long-distance travel, Non-negative tensor factorization (NTF), Mobile-phone location data

*Corresponding author
Email address: hamaguchi@se.kanazawa-u.ac.jp (Hiromichi Yamaguchi)

Highlights

- We discussed the base travel groups for forecasting the new HSR effect.
- We applied the mobile phone location data and NTF for the estimation.
- The optimal number of base travel groups is twelve in the case of Hokuriku HSR.
- The estimated sensitivity to the new HSR varies between the base travel groups.
- The estimated base travel groups are related to travel purposes.

1. Introduction

Numerous long-distance travel demand models have been proposed to evaluate nation-wide transportation projects such as the high-speed rail (HSR). In these models, the respective parameters are estimated for each travel purpose, with different elasticities for certain service changes [1, 2]. It can be presumed that the sensitivity differs among the various travel purposes (*e.g.*, sightseeing, business); therefore, the different parameters are estimated and applied for each travel purpose. Hence, the method of decomposing travel behavior into multiple purposes with different elasticities is important for the accuracy of demand forecasting. Here, the 'base travel group' is defined as a travel group with inherent sensitivity. A typical classification is the travel purpose, but in this study, we focus on the fact that the sensitivity differs among various travel purposes and we do not limit the study to classification based on the travel purpose.

The number of base travel group in the travel demand models has been determined empirically in the most previous studies. Three major travel purpose groups, *i.e.*, business, sightseeing, and private, have been frequently considered for the demand forecasting model. For example, the 'travel purposes' of the national inter-regional travel survey in Japan [3] are defined as the same three groups. The same travel group classification (business, sightseeing, and private) is often used in any other surveys [4, 5]. However, the validity of these base travel classification is rarely discussed.

This is owing to the difficulty in investigating the long-distance travel behavior of the entire country in detail [6], which requires the collection of numerous samples because the behavior occurs at low frequencies and differs depending on the individual. In addition, as Janzen et al. (2018); Madre et al. (2007) [7, 8] have pointed out, questionnaire surveys in which journey histories are queried tend to report an extremely low number of journeys due to high response burden and memory issues. Therefore, to conduct a countrywide survey, it is required that the questions are as simple as possible, which makes it difficult to detail the classification for travel purposes.

This implies that it is difficult to understand the effect of the HSR service on travel behavior, which is expected to have a significant nationwide impact. Therefore, in many studies, the local impact alone of the HSR service in each destination has been analyzed (*e.g.* Gao et al. [9]). Several nationwide travel models are often estimated only a few types of sensitivities in each purpose [1, 2]. However, it is unlikely that the sensitivity to the HSR effect will remain constant

among different travel purposes. For example, the sensitivity of travel behaviors to the new HSR service differs between the summer-vacation season and other seasons, even during the same travel purpose, for example, sightseeing. Moreover, the sensitivity is different among various age groups, depending on the differences in hobbies at the travel destination. Such variations have been demonstrated in many studies that have analyzed the travelers' satisfaction of HSR services (*e.g.* Chou and Yeh (2013); Chou et al. (2014); High et al. (2018) [10, 11, 12]). Although these studies have analyzed many differences in sensitivity to the HSR service, it is difficult to understand the nationwide travel volume variation owing to the small number of samples. In other words, although a large difference in sensitivity is expected for each "travel group" defined by the travel purpose, season, and individual attributes, its observation is difficult by conventional questionnaire surveys.

On the other hand, mobile-phone location data, used in recent years, enable the easy obtainment of data on long-distance travel behavior at arbitrary points. This is because the location data of numerous samples (mobile-phone users) are frequently recorded over a wide range without any response burden and memory issues. However, these data include certain disadvantages; one is that information on the intentions of people such as the travel purpose cannot be obtained. This is inconvenient for creating a model for groups with different elasticities to service changes. Therefore, Alexander et al. (2015); Janzen (2016) [13, 14] proposed methods for estimating the travel purpose by combining a plurality of data.

In this study, we propose an approach to derive the base travel groups with different sensitivities for the new HSR service directly from the mobile-phone location data to overcome the difficulties in observing long-distance travel behavior. In the proposed approach, mobile-phone-location data with time-series change and individual attribute information alone are used. In other words, no additional information is required, unlike in the purpose estimation approach [14, 15]. The proposed approach has the following advantages. First, we can eliminate the limitation of questionnaire surveys, which could handle only a small number of sensitivity groups, owing to their response burden and memory issues. This leads to the creation of a more accurate long-distance travel demand forecasting model, owing to the detailed sensitivity information. Second, we can obtain the sensitivities of several base travel groups, which are defined by the detailed time-series change and individual attribute information. The lack of travel purpose information in the mobile-phone-location data is compensated by a group based on the time-series change and

individual attribute information. Third, the proposed approach is applicable to longer time data, and enables us to analyze longer changes that are stable for several years. In the evaluation of the long-distance travel service, it is desirable to understand effects that last for longer periods, rather than short-term effects, such as increases of travelers immediately after the commencement of new HSR service. By using the proposed approach based on mobile-phone-location data alone, it is easy to obtain and apply data at any time-point.

This approach is the same as mechanically extracting an activity pattern (factorization) from long-term travel data, and has already been applied for travel data in many studies. For example, micro travel history data have been used for analyzing the 'activity-patterns' in many studies, and various models for the analysis have been proposed [16, 17, 18]. In contrast, this study involves the application of passively-recorded data without travel purpose information. Therefore, the travel purpose information needs to be inferred from pattern-mining. Ahas et al. (2008); Yao et al. (2015); Sun and Axhausen (2016); Chen et al. (2019) [19, 20, 21, 22] attempted to mine the travel patterns from a similar dataset, and successfully obtained useful travel behavior patterns using several decomposition methods, such as principal component analysis (PCA). However, the approach proposed by Ahas et al. (2008); Yao et al. (2015); Sun and Axhausen (2016) [19, 20, 21] has the following two limitations, which can be overcome by our proposed approach.

First, the decomposition result contains several negative values. This is inappropriate when discussing the travel volume and the difference in sensitivity against the new HSR service, because it does not satisfy the additive property of the number of travelers. Let us consider the data in Table 1 as an example. Here, there are two zones A and B, two seasons 1 and 2, and two base travel groups with different sensitivities to the new HSR service. The values in the table indicate the number of travelers from zones A and B to zone D at the corresponding time. The two travel groups are assumed to have different sensitivities to HSR and spatiotemporal distributions before the commencement of the new HSR service. After the commencement of the new HSR service (connecting from zones A and B to zone D), the number of travelers varies depending on the sensitivity.

We can only observe the total number of two travel groups from the mobile-phone-location

Table 1: The sample of time-series change of travel volume to zone D where connected by new HSR service

Time	total (Observed)		High sensitivity (Not observed)		Low sensitivity (Not observed)	
	Zone: A	Zone: B	Zone: A	Zone: B	Zone: A	Zone: B
Before commencement of service						
Season 1	12	18	8	2	4	16
Season 2	44	26	40	10	4	16
After commencement of service						
Season 1	21	24	16	4	5	20
Season 2	85	40	80	20	5	20

data. The observed data can be decomposed by factor analysis, as follows:

$$\begin{bmatrix} 12 & 18 \\ 44 & 26 \\ 21 & 24 \\ 85 & 40 \end{bmatrix} - \begin{bmatrix} 40.5 & 27 \\ 40.5 & 27 \\ 40.5 & 27 \\ 40.5 & 27 \end{bmatrix} = \begin{bmatrix} 0.97 & 29.87 \\ 1.91 & -3.10 \\ -2.38 & 19.59 \\ -0.50 & -46.36 \end{bmatrix} \begin{bmatrix} 0.27 & -0.96 \\ -0.97 & -0.27 \end{bmatrix} \tag{1}$$

$$= \begin{bmatrix} 0.97 \\ 1.91 \\ -2.38 \\ -0.50 \end{bmatrix} \begin{bmatrix} 0.27 & -0.96 \end{bmatrix} + \begin{bmatrix} 29.87 \\ -3.10 \\ 19.59 \\ -46.36 \end{bmatrix} \begin{bmatrix} -0.97 & -0.27 \end{bmatrix} \tag{2}$$

Here, the difference between the observed total travel volume and the average value for each zone is expressed as the product of two matrices. There are negative values in one of the matrices on the right-hand-side of the equation. These decompositions enable us to understand the existence of two different trends in the observed data. However, unlike the unobserved data of decomposed groups in Table 1, it is difficult to answer the following questions: 'How many sensitive travelers are present in a certain season/area?' 'What is the sensitivity of each travel group?' This is because the additive property and units of data are not retained after the factor analysis. It is important to answer the above questions quantitatively for the evaluation and prediction of new

HSR services. To answer these questions, the following non-negative decomposition is desirable:

$$\begin{bmatrix} 12 & 18 \\ 44 & 26 \\ 21 & 24 \\ 85 & 40 \end{bmatrix} = \begin{bmatrix} 10 & 20 \\ 50 & 20 \\ 20 & 25 \\ 100 & 25 \end{bmatrix} \begin{bmatrix} 0.80 & 0.20 \\ 0.20 & 0.80 \end{bmatrix} \tag{3}$$

$$= \begin{bmatrix} 10 \\ 50 \\ 20 \\ 100 \end{bmatrix} \begin{bmatrix} 0.80 & 0.20 \end{bmatrix} + \begin{bmatrix} 20 \\ 20 \\ 25 \\ 25 \end{bmatrix} \begin{bmatrix} 0.20 & 0.80 \end{bmatrix} \tag{4}$$

Here, in equation (4), the observed matrix is decomposed into two travel groups represented by the product of the temporal vector (left-hand-side) and the spatial distribution vector (right-hand-side). These matrices directly yield the sensitivity, spatiotemporal distribution, and quantity of each travel group. Therefore, in this research, we propose an approach for directly estimating the unobserved base travel groups in Table 1.

The second limitation is the difficulty in extracting information while focusing on the sensitivity to the new HSR service, even with the additivity property and units being retained. Because long-distance travel behavior varies significantly with time, a large number of time-series components may be detected during the decomposition of the time-series data. In particular, if the new HSR effect is less significant than the seasonal fluctuations, it is difficult to extract the new HSR effect from a large amount of detection results. Therefore, in this research, we propose a method to identify and extract the changes in the number of travelers before and after the commencement of the new HSR service by combining models representing some specific time-series patterns. Furthermore, we propose an efficient algorithm for this non-negative decomposition approach as an extension of Lee and Seung (2001) [23].

The proposed approach is applied to the time-series data of a number of long-distance travelers to the Ishikawa Prefecture, to where the Hokuriku HSR has been newly extended on 15th March 2020. Based on this, the number of base travel groups necessary for predicting the effect of a new HSR is estimated. Furthermore, the characteristics of the change in the travel demand due to the new HSR are clarified.

The remainder of this paper is organized as follows: The Japanese mobile-phone location data are described, and the basic information directly obtained from these data are presented.

Subsequently, the developed decomposition method for extracting a couple of base travel groups is presented. Based on the decomposed results, several features of the base groups derived from the time-series of the long-distance travel data are described.

2. Data: Mobile phone location data

The 'mobile spatial statistics' (MSS) by NTT DOCOMO, INC. provides the population of the residence (location of permanence) and staying area (location of temporality)pairs. Nakanishi et al. (2018) [24] estimated this population through the following procedure using the periodic communication records between mobile phones and base stations, as follows: First, the number of mobile phones for each base station was obtained by aggregating the records. Next, the population was calculated by multiplying by the reciprocal of the share of each attribute, such as the place of residence. These communication records are called "network-driven cellular network-based data" in inWang et al. (2018) [25]. The other type of cellular network-based data is "event-driven data", such as CDRs [7, 26]. Network-driven data are collected periodically [1] without event triggers such as incoming calls; the temporal granularity of network-driven data is higher and more stable than that of event-driven data.

As the number of communication between the mobile terminal network and approximately 50 million mobile phones are recorded for the service operation of NTT DOCOMO, INC., the population in any residence-staying area pair can be obtained at any time using this method. This type of data can provide more comprehensive new insights into travel patterns, *e.g.*, understanding the seasonal variations, and discovering the correlations and time-series changes in the patterns. This section presents the nationwide population distribution data used for examining the effects of the Hokuriku HSR in Japan. The MSS data is the aggregated population distribution data estimated using the operation data of the NTT DOCOMO mobile terminal network. The estimation methodology of the MSS data is presented in Terada et al. (2013) [27]. The aggregated population distribution data sample used in our analysis is presented in Table 2. This study uses the estimated population for each combination of time, detected (staying) zone, residence zone, age, and gender. Here, the detected (staying) zone indicates the location at which each mobile phone is present at a given time. The residence zone, age, and gender information are obtained

[1] In the case of NTT Docomo, the records are collected every hour.

from the contract information. The detected and residence zones are classified in the prefecture, where Japan is divided into 47 zones.

Table 2: The sample of MSS data

Date	time	detected zone	residence zone	age	gender	estimated population
2014/03/01	13:00	Ishikawa	Tokyo	20-29	male	111
2014/03/01	13:00	Ishikawa	Tokyo	20-29	female	111
2014/03/01	13:00	Ishikawa	Tokyo	30-39	male	111
2014/03/01	13:00	Ishikawa	Tokyo	30-39	female	111
2014/03/01	13:00	Ishikawa	Tokyo	40-49	male	111
2014/03/01	13:00	Ishikawa	Tokyo	40-49	female	111

In this study, we analyze the changes in the travel groups caused by the Hokuriku HSR (denoted by the blue line in Figure 1). This new HSR service, which has been operational since May 14, 2015, has reduced the travel time by railway between Tokyo and Kanazawa (main city of the Ishikawa Prefecture) from 230 min to 150 min. As a result, the group of long-distance travel behavior to the Ishikawa Prefecture (green zone in Figure 1) is expected to change significantly.

To analyze the changes in the travel behavior, we analyze the three-way tensor $\mathbf{Y}$ obtained from the MSS. Each component ($y_{d,i,a}$) indicates the number of travelers staying at the Ishikawa Prefecture at 13:00 on day $d \in D$; this number is divided by their residence zone $i \in I$ and individual attribute (age and gender) a. $i \in I$ indicates the residence zone of the traveler. Here, as the spatial-zone, we use the prefecture unit (Japan is divided into 47 prefectures); we use 44 zones excluding Ishikawa, and two adjacent zones (Toyama and Fukui prefectures) as the residence zone set I, for analyzing long-distance travel. Attribute $a \in A$ is defined as a combination of the gender and age. Here, five age groups (20-29, 30-39, 40-49, 50-59, 60-69) are considered in the analysis. Therefore, the attribute set A contains $10 = 2 \times 5$ components.

We analyze the time-series changes for four years, (*i.e.,* 1,461 days, D = [Mar. 1, 2014, ..., Feb. 28, 2018]). Hence, the target three-way tensor $\mathbf{Y}$ has 1,461 (number of days) $\times$ 44 (residence zone) $\times$ 10 (attributes) components.

To understand the basic time-series pattern, we overview the total number of travelers $\tilde{y}_d$ to the Ishikawa prefecture. This value is defined by following equation:

$$\tilde{y}_d = \sum_{i \in I} \sum_{a \in A} y_{d,i,a}. \tag{5}$$

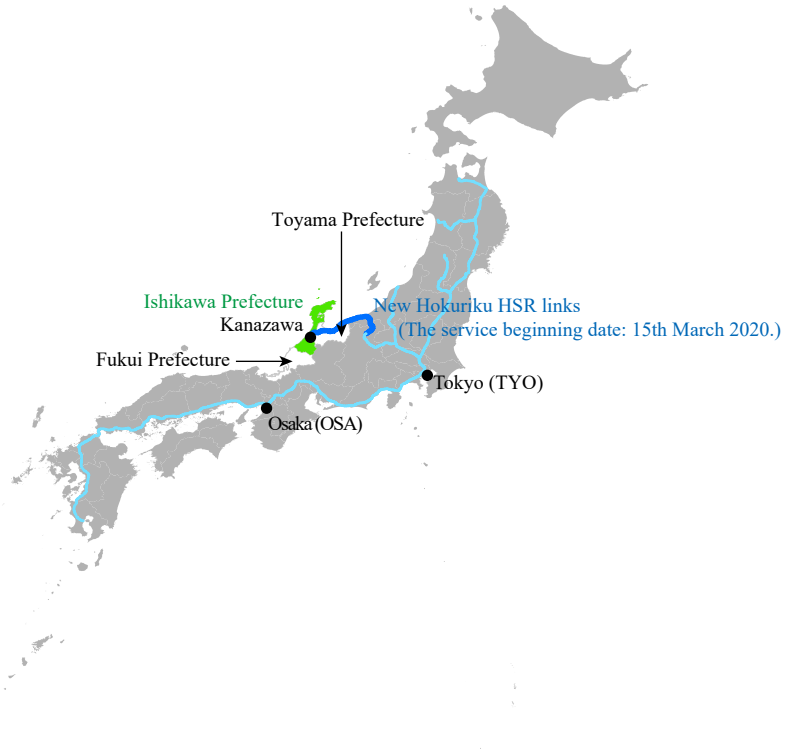


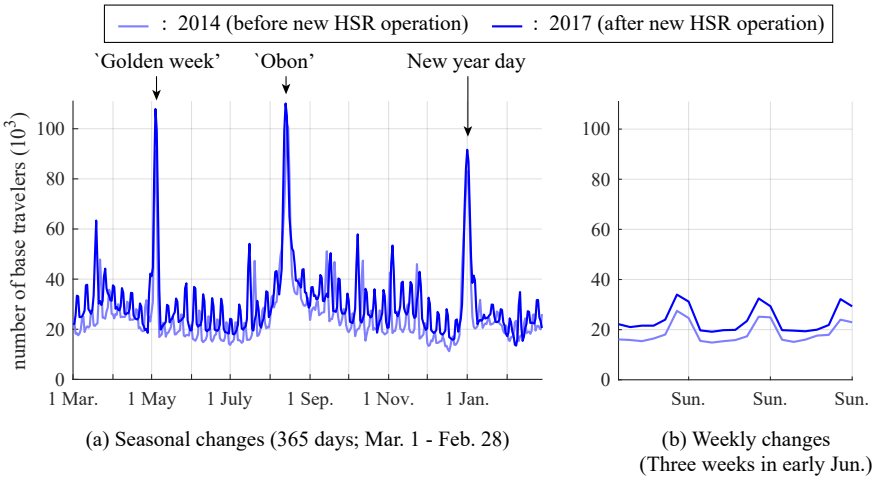
Figure 1: The high-speed rail (HSR) network in Japan and the new Hokuriku high-speed rail links

Figure 2, which displays $\tilde{y}_d$, indicates the following four features of long-distance travelers: 1) There are numerous travelers, particularly during the following three periods: year-end and new year, summer vacation (Obon: religious vacations in Japan), and the golden week (several consecutive national holidays in early May). 2) The number of travelers is more in the spring and summer seasons compared to the others. 3) The time-series change in the 7-day cycle is large, and is higher on weekends compared to weekdays. (see Figure 2 (b)) 4) Overall, the number of travelers was more in 2017 compared to 2014, and this increase includes the effect of the new HSR.

3. Non-negative tensor factorization for time-series data

3.1. NTF model for time-series data

In this study, several travel groups are extracted from tensor $\mathbf{Y}$ through NTF, which is the generalized model from the non-negative matrix factorization [28] to tensor data. In simple NTF



(a) Seasonal changes (365 days; Mar. 1 - Feb. 28)

(b) Weekly changes
(Three weeks in early Jun.)

Figure 2: Time-series change in the total number of travelers $\tilde{y}_d$ to the Ishikawa prefecture

(CP decomposition), tensor $\mathbf{Y} \in \mathbb{R}_{+}^{\#(D)\times\#(I)\times\#(A)}$ is decomposed into the products of three non-negative matrices, $\boldsymbol{U} \in \mathbb{R}_{+}^{\#(D)\times\bar{k}}$, $\boldsymbol{V} \in \mathbb{R}_{+}^{\#(I)\times\bar{k}}$, and $\boldsymbol{W}_{+} \in \mathbb{R}^{\#(A)\times\bar{k}}$, as follows:

$$\mathbf{Y} = \mathbf{Z} + \mathbf{E} = \sum_{k=1}^{\bar{k}} \boldsymbol{u}_k \circ \boldsymbol{v}_k \circ \boldsymbol{w}_k + \mathbf{E}. \tag{6}$$

Here, $\mathbf{Y}$ is the observed tensor of the MSS data, $\mathbf{Z}$ is the tensor composed of three non-negative components, $\circ$ is the outer product, $\#(D)$ is the number of components in set D (here, $\#(D) = 1,461$), and $\mathbf{E} \in \mathbb{R}^{\#(D)\times\#(I)\times\#(A)}$ is the error tensor. Vectors $\{\boldsymbol{u}_k, \boldsymbol{v}_k, \boldsymbol{w}_k\}$ are treated as the column vectors of the factor matrices, such as $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{\bar{k}}]$. Figure 3 is the graphical representation of this CP decomposition. As depicted in this figure, this approach divides the observed tensor $\mathbf{Y}$ into several components (travel groups) and an error tensor $\mathbf{E}$, where $\bar{k}$ is the number of base travel groups.

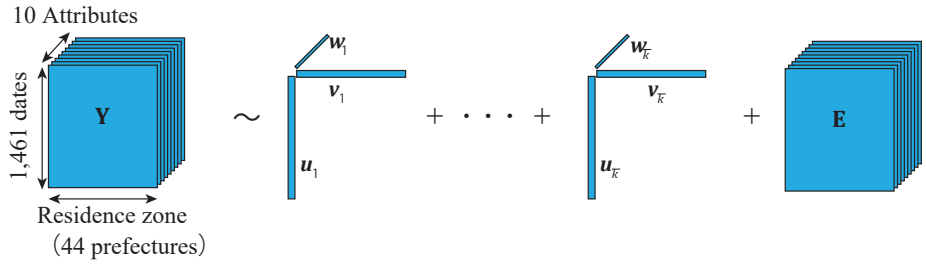


Figure 3: Time-series change in the total number of travelers $\tilde{y}_d$ to the Ishikawa prefecture

In an component-wise form, equation (6) can be expressed as follows:

$$z_{d,i,a} = \sum_{k=1}^{\bar{k}} u_{d,i,k} v_{i,k} w_{a,k}, \tag{7}$$

$$\sum_{i \in I} v_{i,k} = 1, \qquad \forall k \in [1, \ldots, \bar{k}], \tag{8}$$

$$\sum_{a \in A} w_{a,k} = 1, \qquad \forall k \in [1, \ldots, \bar{k}], \tag{9}$$

$$u_{d,i,k} \geq 0, \qquad \forall (d, i, k) \in (D \times I \times [1, \ldots, \bar{k}]), \tag{10}$$

$$v_{i,k} \geq 0, \qquad \forall (i, k) \in (I \times [1, \ldots, \bar{k}]), \tag{11}$$

$$w_{a,k} \geq 0, \qquad \forall (a, k) \in (A \times [1, \ldots, \bar{k}]). \tag{12}$$

Here, equations (10–12) are non-negative constraints which are the major features of the NTF. Vectors $\{\boldsymbol{v}_k, \boldsymbol{w}_k\}$ constituting matrices $\boldsymbol{V}$ and $\boldsymbol{W}$ indicate the travel groups expressed by a configuration of the travel origin and traveler attribute; the sum of components in each vector $\{\boldsymbol{v}_k, \boldsymbol{w}_k\}$ is equal to one, as shown in equations (8, 9). Each component $v_{i,k}, w_{a,k}$ indicates the ratio occupied by the residence zone i or attribute a in base travel group k. Component $u_{d,i,k}$ of matrix $\boldsymbol{U}$ indicates the number of travelers who visited to the Ishikawa prefecture by base travel group k on day d. $\bar{k}$ is the number of base travel groups. It is necessary to determine this number in advance; the method used to obtain this number in this study is described later.

Matrix $\boldsymbol{u}_k \circ \boldsymbol{v}_k$ is further decomposed into periodic patterns and changes before and after the commencement of the new HSR service for extracting the representative temporal changes. Specifically, we employ an approach to explain the time-series changes in the number of travelers, using several periodic (calendar) components and the HSR effect alone. Denoting the periodic components and HSR effect by a dummy variable matrix $\boldsymbol{M}$, equation (7) can be rewritten

as follows:

$$z_{d,i,a} = \sum_{k=1}^{\bar{k}} \alpha_{d,i,k} \left(\sum_{p \in P} m_{d,p} x_{p,k} \right) v_{i,k} w_{a,k}, \tag{13}$$

$$x_{p,k} \geq 0, \qquad \forall (p,k) \in (P \times [1, \ldots, \bar{k}]), \tag{14}$$

$$m_{d,p} \in (0, 0.5, 1), \qquad \forall (d,p) \in (D \times P), \tag{15}$$

$$P = P_{\text{week}} \sqcup P_{\text{year}}, \tag{16}$$

$$P_{\text{week}} = [\text{Mon.}, \text{Tue.}, \ldots, \text{Sun.}]. \tag{17}$$

$$P_{\text{year}} = [\text{Jan.1}, \text{Jan.2}, \ldots, \text{Mar.31}], \tag{18}$$

Here, a set P of periodic components p is composed of two types of cycles, as formulated in equations (16 – 17); $\sqcup$ is the union of the two sets. The first periodic pattern (set P_{week} that includes seven components) represents the time-series changes in a one-week cycle, whereas the second pattern (set P_{year} that include 365 components) represents that in a one-year cycle. Variables $m_{d,p}$ are dummy variables representing the relationship between the date d and periodic component p (1 for the corresponding day and 0 otherwise). The exception is a leap day, where the average value of the components of the previous and following days is applied for $m_{d,p}$ as follows:

$$m_{(d=\text{Feb.29,2016}),p=\text{Feb.28})} = m_{(d=\text{Feb.29,2016},p=\text{Mar.1})} = 0.5. \tag{19}$$

In this model, $\sum_{p \in P} m_{d,p} = 2$ is satisfied because there are two types of periodic components. $x_{p,k}$ is a parameter that directly indicates the number of travelers per day for each periodic component p and base travel group k.

Furthermore, $\alpha_{d,i,k}$ depict the change caused by the HSR in travel group k. These variables are formulated as follows:

$$\alpha_{d,i,k} = \begin{cases} 1 & (\forall d \in D_{\text{bef.}}) \\ 1 + \gamma_{i,k} & (\forall d \in D_{\text{aft.}}) \end{cases} \qquad \forall (i,k) \in (I \times [1, \ldots, \bar{k}]), \tag{20}$$

$$\gamma_{i,k} \geq 0, \qquad \forall k \in [1, \ldots, \bar{k}], \tag{21}$$

$$D_{\text{bef.}} = [\text{Mar.1}, 2014, \ldots, \text{Mar.14}, 2015], \text{(before the HSR extension)}, \tag{22}$$

$$D_{\text{aft.}} = [\text{Mar.15}, 2015, \ldots, \text{Feb.28}, 2018], \text{(after the HSR extension)}. \tag{23}$$

. Parameter $\gamma_{i,k}$ is applied to $\alpha_{d,i,k}$ on the date on which the HSR service commenced ($D_{\text{aft.}}$).

Hence, this parameter represents the change rate of travel group k before and after the commencement of the HSR service. Here, this parameter is defined in each residence zone i because the new HSR effect is expected to be different for each OD pair. For example, a particularly large change is expected in the residence-destination pair to which the new HSR is directly connected. Moreover, equation (21) indicates that this parameter has a constraint wherein it denotes the positive changes alone. This constraint renders it possible to output a common travel group before and after the commencement of the new HSR service. [2]

3.2. Estimation method

Based on the above assumption, unknown matrices $\boldsymbol{\Gamma}$, $\boldsymbol{X}$, $\boldsymbol{V}$, and $\boldsymbol{W}$ are estimated to minimize the Frobenius norm of error tensor $\mathbf{E}$ as follows:

$$[\boldsymbol{\Gamma}^*, \boldsymbol{X}^*, \boldsymbol{V}^*, \boldsymbol{W}^*] = \operatorname{argmin} \{\|\mathbf{Y} - \mathbf{Z}\|_2 \mid \text{equations (8), (9), (11), (12), (14), (20), (21)}\}. \quad (24)$$

Here, tensor $\mathbf{Y}$ represents the input data and the four matrices $\boldsymbol{\Gamma}^*$, $\boldsymbol{X}^*$, $\boldsymbol{V}^*$, and $\boldsymbol{W}^*$ represent the output data for understanding the new HSR effects included in $\mathbf{Y}$.

The iterative algorithm proposed by Lee and Seung (2001) [23] can be applied for this estimation. The algorithm derived from the auxiliary function method and Jensen's inequality can be applied directly to nonnegative tensor decomposition with additional conditions, such as our proposed model. In this approach, the following update formulas are obtained from equation (24) (the detailed are written in Appendix A.):

$$\gamma_{i,k} \leftarrow \frac{\sum_{(d,a)\in(D_{\text{aft.}}\times A)} y_{d,i,a} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) v_{i,k} w_{a,k}}{\sum_{(d,a)\in(D_{\text{aft.}}\times A)} z_{d,i,a} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) v_{i,k} w_{a,k}} (\gamma_{i,k} + 1) - 1, \quad (25)$$

$$x_{p,k} \leftarrow \frac{\sum_{(d,i,a)\in(D\times I\times A)} y_{d,i,a} \alpha_{d,i,k} v_{i,k} w_{a,k} m_{d,p}}{\sum_{(d,i,a)\in(D\times I\times A)} z_{d,i,a} \alpha_{d,i,k} v_{i,k} w_{a,k} m_{d,p}} x_{p,k}. \quad (26)$$

$$v_{i,k} \leftarrow \frac{\sum_{(d,a)\in(D\times A)} y_{d,i,a} \alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) w_{a,k}}{\sum_{(d,a)\in(D\times A)} z_{d,i,a} \alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) w_{a,k}} v_{i,k}. \quad (27)$$

$$w_{a,k} \leftarrow \frac{\sum_{(d,i)\in(D\times I)} y_{d,i,a} \alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) v_{i,k}}{\sum_{(d,i)\in(D\times I)} z_{d,i,a} \alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p} x_{p,k}\right) v_{i,k}} w_{a,k}. \quad (28)$$

[2] When we decompose using a model that does not include this constraint, some of the components are almost zero after the commencement of the HSR service, indicating that the model describes the changes due to the new HSR service by replacing the components. However, this result is inconvenient for estimating the sensitivity of each travel group because it is unlikely that a defined specific travel group will disappear before and after the commencement of the new HSR service. Therefore, it is assumed that the new HSR has no effect in reducing the OD travel volume.

The calculation load for this update is considerable with a large-sized tensor. However, using the partitioned parallel computing algorithm for CUDA (computation framework for GPU) proposed by Antikainen et al. (2011) [29], it can be calculated with sufficient speed.

The solution and convergence provided by this iterative algorithm is highly dependent on the initial conditions as discussed in literature [30]. In our study, the initial value is set by 'multistart initialization' [30]. Specifically, the initial parameters are randomly set 40 times, and the update algorithm expressed by equations 25–28 is executed 20 times with each randomly initialized parameter. Subsequently, the parameter set with the least mean square error (MSE) is adopted, and the update algorithm is executed until convergence.

3.3. Optimal number of base travel groups

The number of travel groups $\overline{k}$ must be determined to perform the abovementioned calculation. Several methods to determine $\overline{k}$ for NMF have been proposed [31, 32]; we use the Bayesian information criterion (BIC) because this model is specifically intended for future HSR effect prediction. The MSE and BIC in this model are calculated using the following equations:

$$\mathrm{MSE} = \frac{\|\mathbf{Y} - \mathbf{Z}\|_2}{\#(\mathbf{Y})} = \frac{\sum_{d\in D}\sum_{i\in I}\sum_{a\in A}(y_{d,i,a} - z_{d,i,a})^2}{\#(\mathbf{Y})}, \tag{29}$$

$$\mathrm{BIC} = \#(\mathbf{Y})\ln(\mathrm{MSE}) + \{\#(\mathbf{\Gamma}) + \#(\boldsymbol{X}) + \#(\boldsymbol{V}) + \#(\boldsymbol{W})\}\ln\{\#(\mathbf{Y})\}, \tag{30}$$

where $\#(\mathbf{Y})$, $\#(\mathbf{\Gamma})$, $\#(\boldsymbol{X})$, $\#(\boldsymbol{V})$ and $\#(\boldsymbol{W})$ are the number of components in tensor $\mathbf{Y}$, $\mathbf{\Gamma}$, $\boldsymbol{X}$, $\boldsymbol{V}$ and matrix $\boldsymbol{W}$, respectively.

Figures 4 and 5 depict the calculated results of the minimum MSE and BIC on varying $\overline{k}$ from 1 to 30. Next, we discuss the relationship between $\overline{k}$ and the BIC based on Figure 4 which indicates that the optimal $\overline{k}$ is twelve. Additionally, we consider the relationship between $\overline{k}$ and the Frobenius norm (MSE) shown in Figure 5. This figure indicates that if $\overline{k} = 12$, the MSE value is 70% lesser that than when $\overline{k} = 1$. It can be observed that even if $\overline{k}$ is increased further, the MSE does not change significantly. Therefore, we set $\overline{k} = 12$ in this study. Moreover, $\overline{k} = 12$ coincides with the value proposed by Hutchins et al. (2008) [32] which is the minimum value of the MSE curve inflection point.

Here, Figure 4 indicates that the demand forecast models (such as that proposed by Fu et al. (2014); Yao and Morikawa (2005) [1, 2]) lack the number of base travel groups (purpose) in BIC. As a result, the time change represented by the three groups exhibits an error approximately 1.6

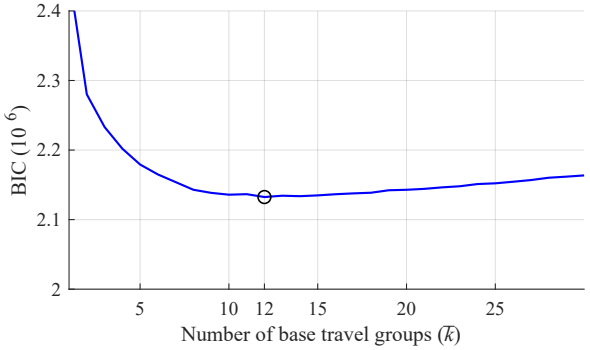


Figure 4: Estimated BIC value

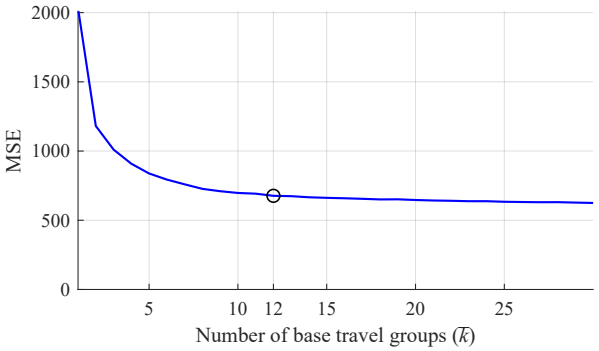


Figure 5: Estimated MSE value

times that obtained in the case of Figure 5. Therefore, we set 12 base travel groups to accurately forecast the demand in this case.

4. Factorization results

4.1. Average travel volume for each travel group

We analyze the results of the twelve decomposed travel groups. Table 3 lists the average daily travel volume for each decomposed group. These values are calculated using the following equations:

$$\overline{u}_{\text{bef.},k} = \frac{\sum_{p \in P_{\text{year}}} x_{p,k}}{365} + \frac{\sum_{p' \in P_{\text{week}}} x_{p',k}}{7}, \tag{31}$$

$$\overline{u}_{\text{aft.},k} = \sum_{i \in I} v_{i,k}(\gamma_{i,k} + 1)\left(\frac{\sum_{p \in P_{\text{year}}} x_{p,k}}{365} + \frac{\sum_{p \in P_{\text{week}}} x_{p,k}}{7}\right). \tag{32}$$

Here, group no.k is assigned a number in descending order, before and after the commencement of the HSR service. Table 3 indicates two features of the decomposed travel groups. The first feature is that the changes in the travel volume before and after the commencement of the HSR service are different among the twelve travel groups. For example, the travel volume of group $k = 12$ has almost no change, whereas those of three other groups increase more than 1.5 times. In our model, the observed change of 1.21 times is represented as a summary of these travel groups with different sensitivities.

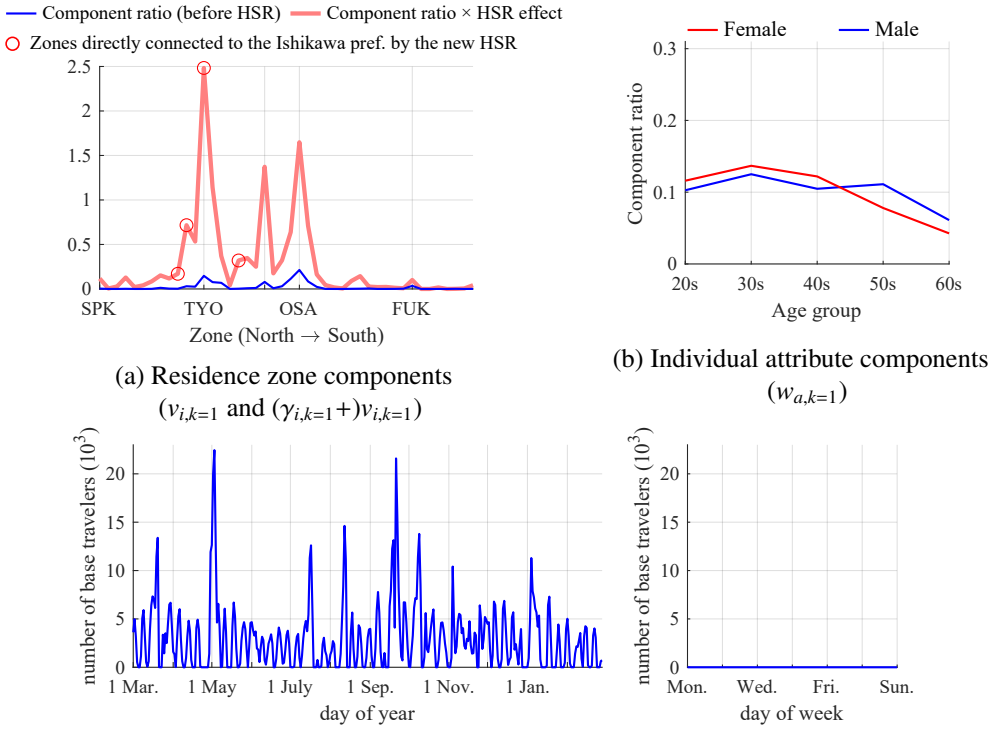
The second feature is that group $k = 1$ alone differs considerably. The volume of travel group $k = 1$ are very different between before and after the commencement of the new HSR service (between $\overline{u}_{\text{bef.},k=1}$ and $\overline{u}_{\text{aft.},k=1}$). Moreover, the volume increases 12 times from before to after. These features indicate that group $k = 1$ is a new travel group created by the commencement of the new HSR service.

Table 3: Number of travelers per day in each decomposed group

group no.(k)	Average (before) $\overline{u}_{\text{bef.},k}$	Average (after) $\overline{u}_{\text{aft.},k}$	Change ratio $\overline{u}_{\text{aft.},k}/\overline{u}_{\text{bef.},k}$
1	236.2	2,989.6	12.655
2	2,433.2	3,897.2	1.602
3	1,689.3	2,680.6	1.591
4	1,014.5	1,423.1	1.401
5	1,492.3	1,988.8	1.336
6	4,524.3	4,740.7	1.046
7	3,145.0	3,264.3	1.042
8	2,624.8	2,707.0	1.033
9	1,826.9	1,871.9	1.025
10	1,994.3	2,001.2	1.004
11	2,235.2	2,236.8	1.002
12	1,998.6	1,995.4	1.000
Total	25,214.7	31,796.6	1.261
Observed	25,430.8	30,770.0	1.210

4.2. Features of the base travel groups

In this subsection, four groups with different new HSR effects are selected and their features are discussed. In addition, considering all the travel groups, the characteristics of those that have changed significantly due to the new HSR are described.



(a) Residence zone components ($v_{i,k=1}$ and $(\gamma_{i,k=1}+)v_{i,k=1}$)

(b) Individual attribute components ($w_{a,k=1}$)

(c) Time-series components ($x_{p,k=1}$)

Figure 6: Residence zone and attribute components of travel group $k = 1$

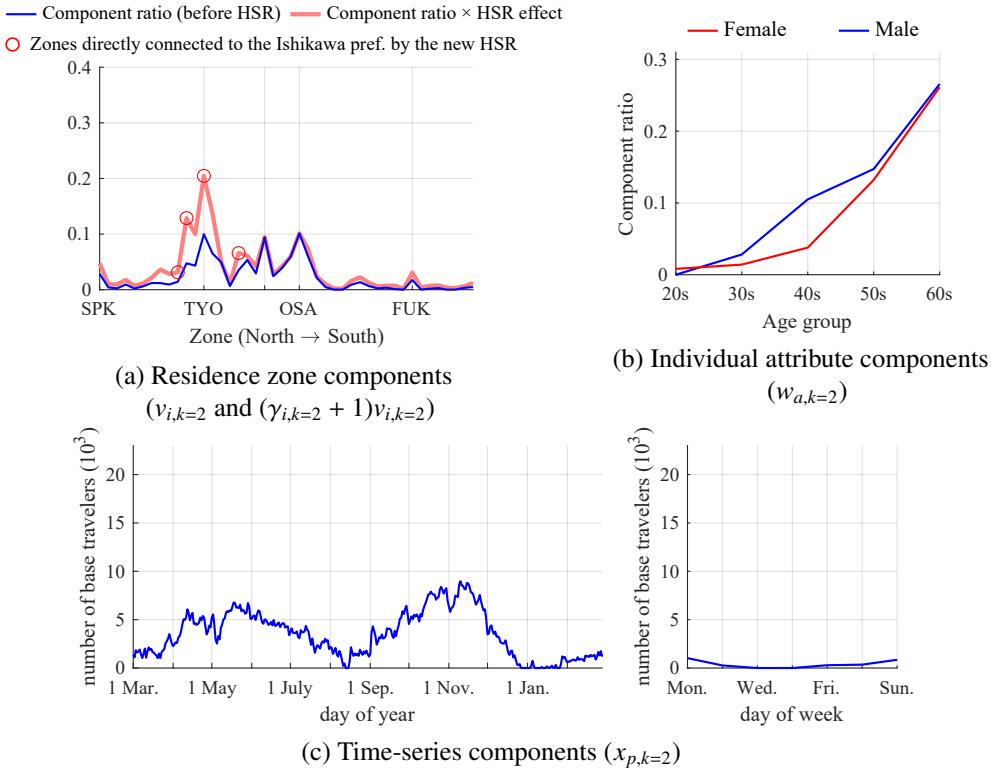
4.2.1. Base group no.1

Let us consider base group no.1. This is a new travel group created by the commencement of the new HSR service, as shown in Table 3. Figure 6 (a) depicts the residence composition, and change before and after the commencement of the new HSR service, for group no.1. In this figure, we can find an expected increase in the number of trips from TYO (Tokyo) to Ishikawa, which are directly connected by the new HSR. A distinct change due to the new HSR effect is also indicated in this travel group; its peculiarity is that there is a considerable increase in travel from places that are not directly connected by the new HSR. A typical example is the increase in travel from the OSA (Osaka) zone. As shown in Figure 1, the Hokuriku HSR is not generally used for travel from the Osaka (OSA) to Ishikawa zones because another direct railway service arrives within a shorter duration. Hence, almost no time-saving effect is expected by the new HSR between the Osaka (OSA) and Ishikawa zones. Nevertheless, in base group no.1, the volume of

travel from Osaka and the other western zones unrelated to the new HSR, also increases.

Figures 6 (b) and (c) depict the attribute composition and time change in total travel volume of this base group, respectively. The attribute composition indicates that base group no.1 includes all the attributes almost equally. The time-series change does not exhibit any weekly cycle trend, but occurs at several points in the yearly cycle.

4.2.2. Base group no.2



(a) Residence zone components ($v_{i,k=2}$ and $(\gamma_{i,k=2}+1)v_{i,k=2}$)

(b) Individual attribute components ($w_{a,k=2}$)

(c) Time-series components ($x_{p,k=2}$)

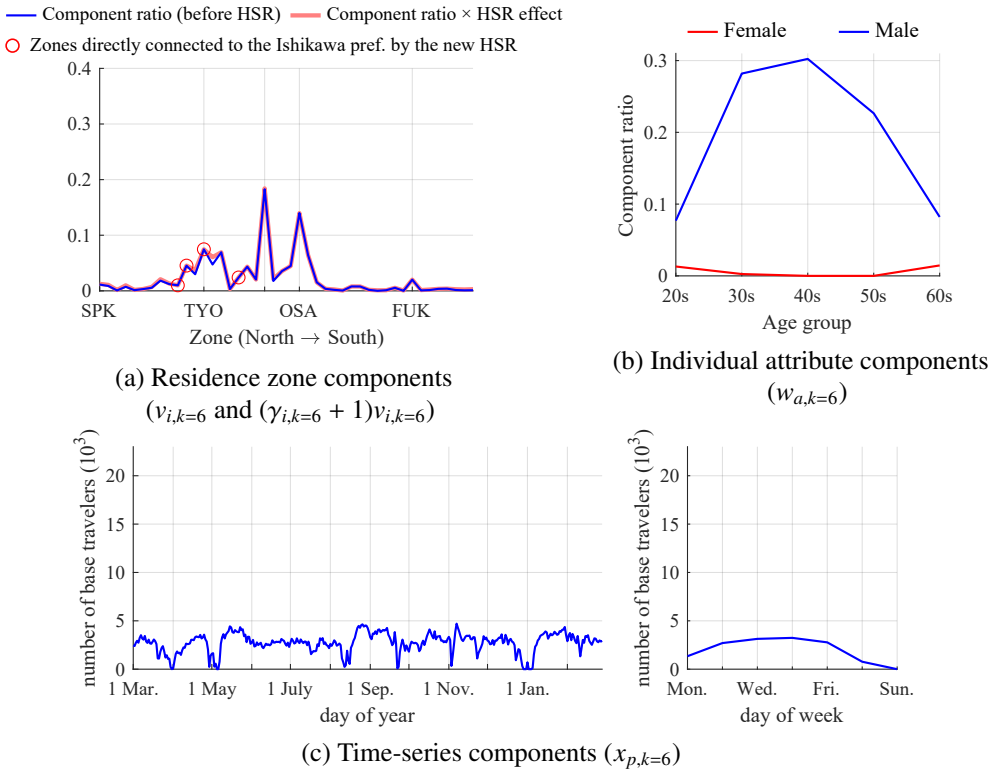
Figure 7: Residence zone and attribute components of travel group $k = 2$

Let us consider base group no.2. Figure 7 (a) shows the increase in the volume of travel by group no.2 in zones directly connected by the new HSR. These change are the expected travel-time saving effects of the new HSR service.

The attribute composition of group no.2, depicted in 7 (b), is mostly occupied by elder people. The periodic time-series change of group no.2, depicted in 7 (c), is considerable in spring (Apr. —Jun.) and autumn (Oct.–Nov.), and rare in winter (Jan.–Mar.). The change in the weekly cycle

is almost zero. Based on the above, this group can be summarized as the travel group of elderly people who visit the Ishikawa prefecture during periods other than the summer vacation and snow season. This travel group is highly sensitive to the travel time reduction by the new HSR service.

4.2.3. Base group no.6



(a) Residence zone components ($v_{i,k=6}$ and $(\gamma_{i,k=6}+1)v_{i,k=6}$)

(b) Individual attribute components ($w_{a,k=6}$)

(c) Time-series components ($x_{p,k=6}$)

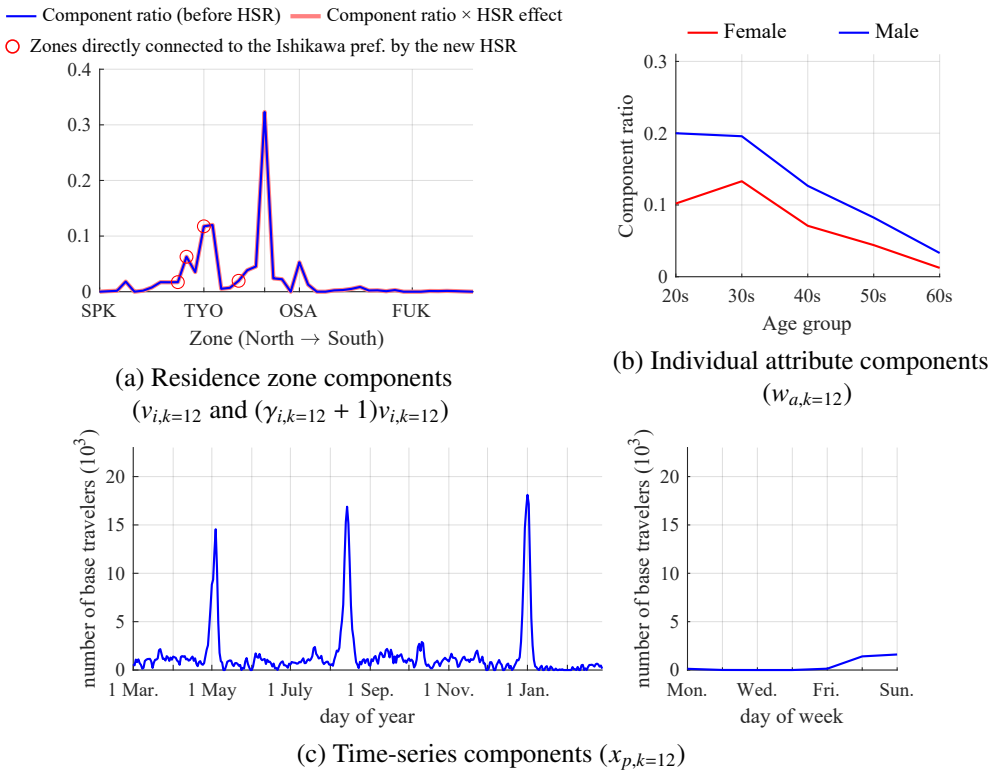
Figure 8: Residence zone and attribute components of travel group $k = 6$

Travel group no.6 does not exhibit considerable change before and after the commencement of the new HSR service, unlike the above two groups, as shown in Figure 8 (a).

Figures 8 (b–c) indicate three features of the attribute component and the periodic time-series change in travel group no.6. The first feature is that this travel group is occupied by 30—50-year old men. The second feature is that the travel volume in weekdays is more than that during weekends. The third feature is that the travel volume of the one-year cycle is almost stable during all the seasons, with certain exceptions. The exception periods match the three special

seasons (Golden-week, Obon, and year-end and new year days) and consecutive holidays due to national holidays. During these days, the volume of travel by group no.6 is less. Based on the above three characteristics, it can be inferred that this travel group has a strong relationship with business-purpose factors, with relatively low-sensitivity to the travel time reduction by the new HSR service.

4.2.4. Base group no.12



(a) Residence zone components ($v_{i,k=12}$ and $(\gamma_{i,k=12}+1)v_{i,k=12}$)

(b) Individual attribute components ($w_{a,k=12}$)

(c) Time-series components ($x_{p,k=12}$)

Figure 9: Residence zone and attribute components of travel group $k = 12$

Travel group no.12 also does not change considerably before and after the commencement of the new HSR service, as shown in Figure 9 (a).

Figures 9 (b–c) indicate two features of the attribute component and periodic time-series change in travel group no.12. The first feature is that this travel group is occupied by younger people. The second is that there are three major peaks in the annual cycle. These three peaks occur are during the Golden-week, Obon, and year-end and new year days. The travel volume

by group no.12 during these periods is more than five times those in the other periods.

In these three seasons, many Japanese return to their hometowns for religious and customary reasons. It can be inferred that travel group no.12 represents the above-mentioned travel behavior to the hometown. This travel group has almost no sensitivity to the travel time reduction by the new HSR service. In other words, travel behavior to hometown at three peaks may be extremely less sensitive to travel time.

4.2.5. Summary of all the base groups

Table 4: Main indicators of all the travel groups

group no.	Change ratio	Time-series rate (/average)		Attribute component ratio	
	$\overline{u}_{\text{aft.},k}/\overline{u}_{\text{bef.},k}$	Weekday	Three seasons [2]	Over 50 years old	Male
1	12.66	1.07	2.33	0.29	0.50
2	1.60	0.98	0.53	0.81	0.55
3	1.59	0.72	1.75	0.45	0.42
4	1.40	1.12	0.90	0.23	0.90
5	1.34	0.75	2.28	0.21	0.47
6	1.05	1.15	0.57	0.32	0.97
7	1.04	0.61	2.21	0.33	0.50
8	1.03	0.84	0.80	0.39	0.66
9	1.02	0.98	5.49	0.17	0.42
10	1.00	0.98	2.75	0.61	0.76
11	1.00	0.88	5.78	0.15	0.45
12	1.00	0.77	5.79	0.17	0.64

Table 4 lists the five indicators of all the decomposed travel groups. The figures depicting detailed information on the other estimated base travel groups are presented in Appendix B. These indicators are selected for depicting the features of the above-mentioned four travel groups. This table indicates the following three points regarding the effect of the new HSR service on the travel groups.

First, the base travel groups no.9, 10, 11, and 12, where travel is concentrated in three special seasons, are less sensitive to the new HSR service. The travel volume of these base travel groups during these three special seasons (Golden-week, Obon, and new-year days) is more than 2.5 times that at other times. The amount of changes in the number of travelers before and after the commencement of the new HSR service is generally less than 1%. Thus, these results indicate that base travel groups, where travel is concentrated in special seasons, are extremely inelastic to

new HSR services.

Second, the travel groups no.4 and 6, in which males travel on weekdays, appear to denote business purposes. While in group no.6, the change before and after the commencement of the new HSR service is small, the change in group no.4 increases approximately 1.4 times. Thus, the groups that travel on weekdays can be decomposed into two groups: one (no.4) with high sensitivity to the new HSR and another (no.6) with low sensitivity.

The third point is that the change ratios are different among the other groups (no.2, 3, 4, 7, 8). these groups appear to represent the sightseeing purpose because the volume during weekdays is less; the increase ratios before and after the commencement of the new HSR service vary from 1.03–1.60, depending on the attributes and time characteristics. The differing sensitivity indicates the necessity for decomposing and extracting these travel groups with different sensitivities using the proposed approach.

5. Conclusions

This study proposed a methodology for deriving the base travel groups necessary for clarifying the time-series changes of the spatial distribution of travelers, applying NTF. This methodology was applied to the time-series data of long-distance travel to the Ishikawa prefecture, where a new Hokuriku HSR service was initiated. The following three conclusions can be drawn from this study: 1) Based on the BIC criteria, the optimum number of base travel groups is twelve. This number is larger than the number of models in previous studies (*e.g.*, [4, 5] use only two models). These detailed sensitivities enable us to make highly accurate forecasts. 2) These groups include components that appear to correspond to different travel purposes (*e.g.*, sightseeing, business, and homecoming), as in previous surveys. These results confirm that the methodology proposed in this study can clearly extract groups with different elasticities, due to the HSR service. In addition, as we had determined similar information on the travel purposes through the proposed methodology, this methodology can be utilized for estimating the travel purpose in detail. 3) The HSR effect can be clarified by dividing it into several characteristics and detailed components.

These estimated groups and their sensitivities to the new HSR service are essential not only for accurate forecasting but also for maximizing the positive effects of the new HSR service in some base travel groups. For example, this information on sensitivities enables us to prepare the

[2]Three special seasons of the Golden week, Obon (summer vacation), and new year holidays.

facilities for traveler groups with higher sensitivity, thereby maximizing the effects of the new HSR. Moreover, this research indicates that we need to focus on the number of travelers traveling from areas that are not connected by the new HSR service, as seen in travel group no.1 in Figure 6. Newly connected areas may increase the number of travelers not only from areas connected by the HSR service but also from other areas. This implies that the launch of the new HSR service is an important opportunity for regions newly connected to the HSR service network, which is required for maximizing its effect to prepare facilities, including other related travel services and advertisements.

In addition, the travel groups obtained in this research enable us to identify the areas that require new HSR services, because new HSR services have a significant impact on travel behavior and individual attributes in certain seasons. In these areas, a considerable amount of long-distance travel is restricted in certain groups owing to the lack of HSR services.

These sensitivity results are estimated from only one addition of the new HSR service. This means that these sensitivities may not be spatially transferred. Therefore, it is required to analyze multiple HSR effects for a more accurate demand-forecasting model in future study. Multiple new HSR effects can be easily analyzed with the proposed model provided the spatial and temporal data of the target range is available.

Most of the interpretations of the base travel groups in this study are supposition. This is the limit of mobile phone location data without information on intentions like travel purposes. On the other hand, prediction of travel volume is reliable, because the traffic volumes for each travel group can be estimated by using similar types of time-series data.

Based on the above research results, we make the following two suggestions to the policy makers. 1) When you are planning a new HSR service, please keep in mind that the new HSR service does not increase all travels. In the case of Japan, long-distance travels by young people during the 3 seasons (Golden week, Obon (summer vacation), Year-end and New Year holidays) did not increase, even between before and after the commencement of new HSR service. Based on this evidence, you need to be careful not to overestimate the effect of the new HSR. 2) Once you have decided to build the HSR, you need to prepare the capacity of the relevant facilities for long-distance travelers. The new HSR only sharply increases travelers for certain seasons and attributes, as seen in figure 6. Furthermore, it may increase the number of travelers from regions not related to new HSR. You need to be prepared to handle these increasing numbers of travelers.

Acknowledgements

The authors would like to thank Mr. Hiroyuki Miyake and Mr. Yuki Kimura for their support in data processing. This work was supported by the JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 18H01560, 18H01556, Young Scientists (B) 15H04060, MEXT LEADER project, and MLIT Committee on Advanced Road Technology (CART).

References

[1] X. Fu, T. H. Oum, J. Yan, An analysis of travel demand in Japan's intercity market: Empirical estimation and policy simulation, Journal of Transport Economics and Policy 48 (PART 1) (2014) 97–113.

[2] E. Yao, T. Morikawa, A study of an integrated intercity travel demand model, Transportation Research Part A: Policy and Practice 39 (4 SPEC. ISSS.) (2005) 367–381. doi:10.1016/j.tra.2004.12.003.
URL https://www.sciencedirect.com/science/article/pii/S0965856404001168

[3] MLIT, 2010 Inter-Regional Travel Survey in Japan, Tech. rep. (2011).
URL http://www.mlit.go.jp/common/001005633.pdf,(lastaccess:2020/05/04)

[4] E. Cascetta, P. Coppola, High Speed Rail (HSR) Induced Demand Models, Procedia - Social and Behavioral Sciences 111 (2014) 147–156. doi:10.1016/J.SBSPRO.2014.01.047.
URL https://www.sciencedirect.com/science/article/pii/S1877042814000482

[5] M. C. Gelhausen, P. Berster, D. Wilken, A new direct demand model of long-term forecasting air passengers and air transport movements at German airports, Journal of Air Transport Management 71 (2018) 140–152. doi:10.1016/J.JAIRTRAMAN.2018.04.001.
URL https://www.sciencedirect.com/science/article/pii/S0969699718301376

[6] K. W. Axhausen, Capturing long-distance travel, Research Studies Press, 2003.

[7] M. Janzen, M. Vanhoof, Z. Smoreda, K. W. Axhausen, Closer to the total? Long-distance travel of French mobile phone users, Travel Behaviour and Society 11 (December 2017) (2018) 31–42. doi:10.1016/j.tbs.2017.12.001.
URL https://doi.org/10.1016/j.tbs.2017.12.001

[8] J.-L. Madre, K. W. Axhausen, W. Brög, Immobility in travel diary surveys, Transportation 34 (1) (2007) 107–128. doi:10.1007/s11116-006-9105-5.
URL http://link.springer.com/10.1007/s11116-006-9105-5

[9] Y. Gao, W. Su, K. Wang, Does high-speed rail boost tourism growth? New evidence from China, Tourism Management 72 (2019) 220–231. doi:10.1016/j.tourman.2018.12.003.

[10] J. S. Chou, C. P. Yeh, Influential constructs, mediating effects, and moderating effects on operations performance of high speed rail from passenger perspective, Transport Policy 30 (2013) 207–219. doi:10.1016/j.tranpol.2013.09.014.

[11] P.-F. Chou, C.-S. Lu, Y.-H. Chang, Effects of service quality and customer satisfaction on customer loyalty in high-speed rail services in Taiwan, Transportmetrica A: Transport Science 10 (10) (2014) 917–945. doi:10.1080/23249935.2014.915247.
URL http://www.tandfonline.com/doi/abs/10.1080/23249935.2014.915247

[12] S.-N. High, F. Zhen, J. Cao, J. Tang, Exploring correlates of passenger satisfaction and service improvement priorities of the, Journal of Transport and Land Use 11 (1) (2018) 559–573. doi:10.5198/jtlu.2018.958.
URL http://dx.doi.org/10.5198/jtlu.2018.958

[13] L. Alexander, S. Jiang, M. Murga, M. C. González, Origin-destination trips by purpose and time of day inferred from mobile phone data, Transportation Research Part C: Emerging Technologies 58 (2015) 240–250. doi:10.1016/j.trc.2015.02.018.
URL http://dx.doi.org/10.1016/j.trc.2015.02.018

[14] M. Janzen, Estimating long-distance travel demand with mobile phone billing data, in: 16th Swiss Transport Research Conference (STRC 2016), Swiss Transport Research Conference (STRC), 2016.
URL http://e-citations.ethbib.ethz.ch/view/pub:178714

[15] A. Suzuki, H. Yamaguchi, D. Fukuda, Data-Fusion Approach for Trip-Purpose Estimation of Inter-Regional Passengers, Proceedings of the 22nd International Conference of Hong Kong Society for Transportation Studies (HKSTS) (2017) 327–338.

[16] R. Kitamura, C. Chen, R. M. Pendyala, R. Narayanan, Micro-simulation of daily activity-travel patterns for travel demand forecasting, Transportation 27 (1) (2000) 25–51. doi:10.1023/A:1005259324588.

[17] A. R. Kuppam, R. M. Pendyala, A structural equations analysis of commuters' activity and travel patterns, Transportation 28 (1) (2001) 33–54. doi:10.1023/A:1005253813277.
[18] C.-H. Joh, T. Arentze, H. Timmermans, Activity-Travel Scheduling and Rescheduling Decision Processes: Empirical Estimation of Aurora Model, Transportation Research Record: Journal of the Transportation Research Board 1898 (1) (2004) 10–18. doi:10.3141/1898-02.
URL http://journals.sagepub.com/doi/10.3141/1898-02
[19] R. Ahas, A. Aasa, A. Roose, Ü. Mark, S. Silm, Evaluating passive mobile positioning data for tourism surveys: An Estonian case study, Tourism Management 29 (3) (2008) 469–486. doi:10.1016/j.tourman.2007.05.014.
[20] D. Yao, C. Yu, H. Jin, Q. Ding, Human mobility synthesis using matrix and tensor factorizations, Information Fusion 23 (2015) 25–32. doi:10.1016/j.inffus.2014.05.005.
URL https://www.sciencedirect.com/science/article/pii/S1566253514000682
[21] L. Sun, K. W. Axhausen, Understanding urban mobility patterns with a probabilistic tensor factorization framework, Transportation Research Part B: Methodological 91 (2016) 511–524. doi:10.1016/J.TRB.2016.06.011.
URL https://www.sciencedirect.com/science/article/pii/S0191261516300261
[22] X. Chen, L. Zhou, Z. Bai, Y. Yue, B. Guo, H. Zhou, Data-Driven Approaches to Mining Passenger Travel Patterns: "left-Behinds" in a Congested Urban Rail Transit Network, Journal of Advanced Transportation 2019 (2019). doi:10.1155/2019/6830450.
[23] D. Lee, H. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Information Processing Systems 13 - Proceedings of the 2000 Conference, NIPS 2000, Neural information processing systems foundation, 2001.
[24] W. Nakanishi, H. Yamaguchi, D. Fukuda, Feature Extraction of Inter-Region Travel Pattern Using Random Matrix Theory and Mobile Phone Location Data, Transportation Research Procedia 34 (2018) 115–122. doi:10.1016/J.TRPRO.2018.11.022.
URL https://www.sciencedirect.com/science/article/pii/S2352146518303119
[25] Z. Wang, S. Y. He, Y. Leung, Applying mobile phone data to travel behaviour research: A literature review, Travel Behaviour and Society 11 (2018) 141–155. doi:10.1016/J.TBS.2017.02.005.
URL https://www.sciencedirect.com/science/article/pii/S2214367X17300224
[26] M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns, Nature 453 (2008) 779–782. doi:10.1038/nature06958.
[27] M. Terada, T. Nagata, M. Kobayashi, Population Estimation Technology for Mobile Spatial Statistics, Docomo Technical Journal 14 (3) (2013) 10–15.
URL https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/rd/technical_journal/bn/vol14_3/vol14_3_010en.pdf,(lastaccess:2020/05/04)
[28] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791. doi:10.1038/44565.
URL http://www.nature.com/articles/44565
[29] J. Antikainen, J. Havel, R. Josth, A. Herout, P. Zemcik, M. Hauta-Kasari, Nonnegative tensor factorization accelerated using GPGPU, IEEE Transactions on Parallel and Distributed Systems 22 (7) (2011) 1135–1141. doi:10.1109/TPDS.2010.194.
[30] A. Cichocki, R. Zdunek, A. H. Phan, S.-i. Amari, Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation, John Wiley & Sons, 2009.
[31] J.-P. Brunet, P. Tamayo, T. R. Golub, J. P. Mesirov, Metagenes and molecular pattern discovery using matrix factorization, Proceedings of the national academy of sciences 101 (12) (2004) 4164–4169.
[32] L. N. Hutchins, S. M. Murphy, P. Singh, J. H. Graber, Position-dependent motif characterization using non-negative matrix factorization., Bioinformatics (Oxford, England) 24 (23) (2008) 2684–90. doi:10.1093/bioinformatics/btn526.
URL http://www.ncbi.nlm.nih.gov/pubmed/18852176http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2639279

Appendix A. Derivation of NTF algorithm

Here, the derivations of update formulas (25) - (28) for solving the optimization problem (equation (24)) are shown. Basically, this process is similar to the iterative algorithm proposed by Lee and Seung (2001) [23].

The objective function $f(\mathbf{\Gamma}, \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{W})$ in the problem is as follows:

$$f(\mathbf{\Gamma}, \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{W}) = \|\mathbf{Y} - \mathbf{Z}\|_2 \tag{A.1}$$

$$= \sum_{(d,i,a)\in(D\times I\times A)} (y_{d,i,a} - z_{d,i,a})^2 \tag{A.2}$$

$$= \sum_{(d,i,a)\in(D\times I\times A)} \left\{ y_{d,i,a}^2 - 2y_{d,i,a}z_{d,i,a} + z_{d,i,a}^2 \right\}. \tag{A.3}$$

When we maximize this function, the difficulty of calculation is in the third term ($z_{d,i,a}^2$), because we need to consider all combinations of terms in $z_{d,i,a}$(equation (13)). Therefore, we use Jensen's inequality and auxiliary function method to approximate this term with a simpler function. Jensen's inequality of convex function h is the theorem formulated as follows:

$$h\left(\sum_i \lambda_i z_i\right) \leq \sum_i \lambda_i h(z_i). \tag{A.4}$$

Based on this inequality, the following inequality is obtained:

$$z_{d,i,a}^2 = \left\{ \sum_{k=1}^{\bar{k}} \left(\alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p}x_{p,k} \right) v_{i,k}w_{a,k} \right) \right\}^2, \tag{A.5}$$

$$\leq \sum_{k=1}^{\bar{k}} \sum_{p\in P} \frac{\alpha_{d,i,k}^2 m_{d,p}^2 x_{p,k}^2 v_{i,k}^2 w_{a,k}^2}{\lambda_{d,i,a,k,p}}, \qquad \forall (d,i,a) \in D\times I\times A, \tag{A.6}$$

$$\lambda_{d,i,a,k,p} \geq 0, \qquad \forall (d,i,a,k,p) \in D\times I\times A\times [1,\cdots,\bar{k}]\times P, \tag{A.7}$$

$$\sum_{k,p\in[1,\cdots,\bar{k}]\times P} \lambda_{d,i,a,k,p} = 1, \qquad \forall (d,i,a) \in D\times I\times A. \tag{A.8}$$

Here, we define function g as follows:

$$g(\mathbf{\Theta}, \boldsymbol{\lambda}) = y_{d,i,a}^2 - 2y_{d,i,a} \sum_{k=1}^{\bar{k}} \left(\alpha_{d,i,k} \left(\sum_{p\in P} m_{d,p}x_{p,k} \right) v_{i,k}w_{a,k} \right) + \sum_{k=1}^{\bar{k}} \sum_{p\in P} \frac{\alpha_{d,i,k}^2 m_{d,p}^2 x_{p,k}^2 v_{i,k}^2 w_{a,k}^2}{\lambda_{d,i,a,k,p}} \tag{A.9}$$

$$\mathbf{\Theta} = [\mathbf{\Gamma}, \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{W}]. \tag{A.10}$$

This function g satisfies the equation (A.11).

$$f(\mathbf{\Theta}) \leq \min_{\boldsymbol{\lambda}} g(\mathbf{\Theta}, \boldsymbol{\lambda}) \tag{A.11}$$

The optimal parameters $\mathbf{\Theta}^* = \mathrm{argmin}_{\mathbf{\Theta}} f(\mathbf{\Theta})$ can be estimated by an algorithm that alternately

and repeatedly updates by following two formulas (auxiliary function method).

$$\boldsymbol{\lambda} \leftarrow \arg\min_{\boldsymbol{\lambda}} \{g(\boldsymbol{\Theta}, \boldsymbol{\lambda}) \mid \text{equations (A.7), (A.8)}\}, \tag{A.12}$$

$$\theta_l \leftarrow \arg\min_{\theta_l} \{g(\boldsymbol{\Theta}, \boldsymbol{\lambda}) \mid \text{equations (8-12), (20-21)}\}, \quad \forall l \in L. \tag{A.13}$$

Here, θ_l is one of the unknown variables $[\boldsymbol{\Gamma}, \boldsymbol{X}, \boldsymbol{V}, \boldsymbol{W}]$. From formula (A.12), the following equation can be obtained analytically.

$$\lambda^*_{d,i,a,k,p} = \frac{\alpha_{d,i,k} m_{d,p} x_{p,k} v_{i,k} w_{a,k}}{\sum_{k=1}^{\bar{k}} \sum_{p \in P} \alpha_{d,i,k} m_{d,p} x_{p,k} v_{i,k} w_{a,k}}, \quad \forall (d, i, a, k, p) \in D \times I \times A \times [1, \cdots, \bar{k}] \times P. \tag{A.14}$$

The update formula of $v_{i,k}$ can be obtained from formula (A.13) as follows:

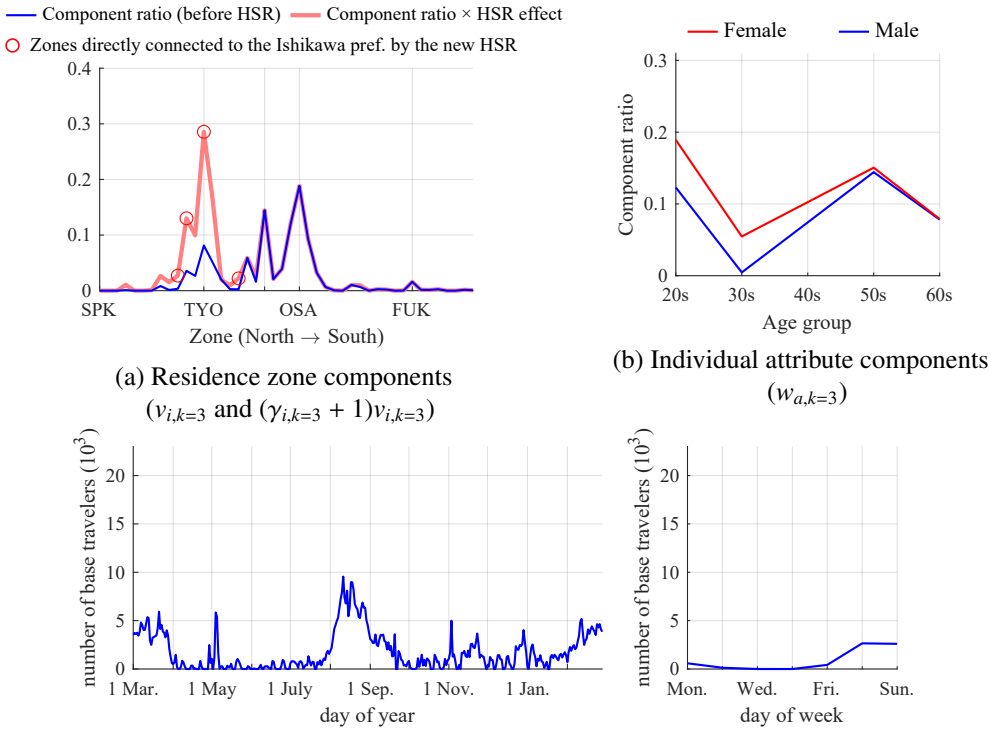
$$\frac{\partial}{\partial v_{i,k}} g(\boldsymbol{\Theta}, \boldsymbol{\lambda}) = 0, \tag{A.15}$$

$$v^*_{i,k} = \frac{\sum_{(d,a,p) \in (D \times A \times P)} \alpha_{d,i,k} m_{d,p} x_{p,k} w_{a,k}}{\sum_{(d,a,p) \in (D \times A \times P)} \frac{\alpha_{d,i,k} m_{d,p} x_{p,k} w_{a,k}}{\lambda_{d,i,a,k,p}}}. \tag{A.16}$$

Then, we can obtain the update formula (27) by substituting equation (A.14) for equation (A.16). The other update formulas (25) - (28) can be derived by the same procedure.

Appendix B. Decomposed results of other travel groups

Appendix B.1. Decomposed travel group no.3



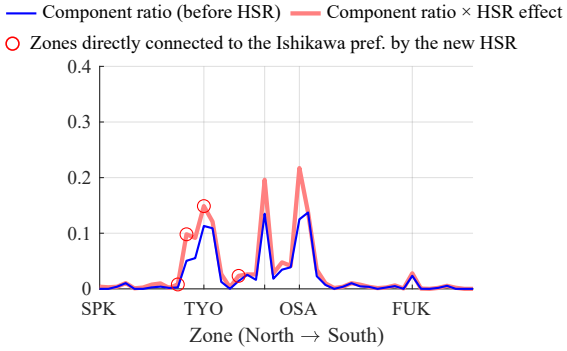
(a) Residence zone components ($v_{i,k=3}$ and $(\gamma_{i,k=3}+1)v_{i,k=3}$)

(b) Individual attribute components ($w_{a,k=3}$)

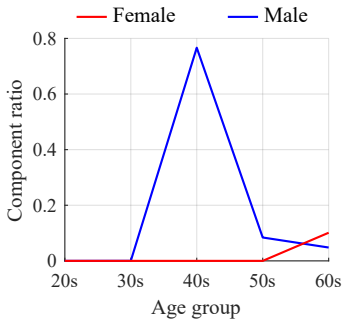
(c) Time-series components ($x_{p,k=3}$)

Figure B.10: Residence zone and attribute components of travel group $k = 3$

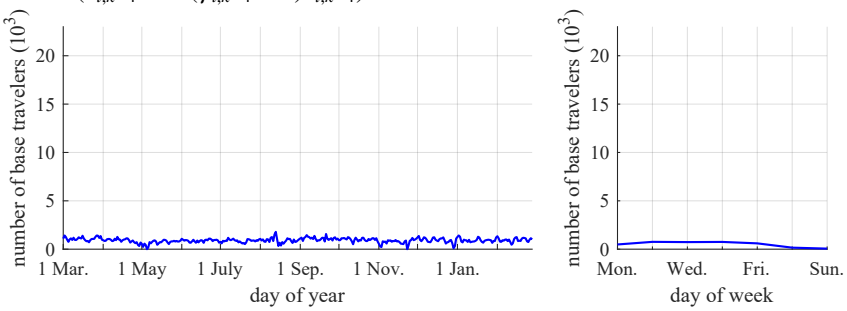
Appendix B.2. Decomposed travel group no.4



(a) Residence zone components ($v_{i,k=4}$ and $(\gamma_{i,k=4}+1)v_{i,k=4}$)



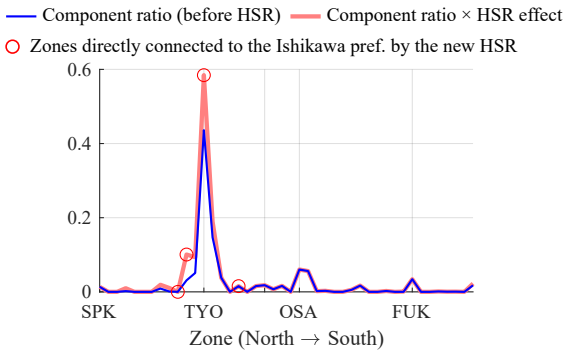
(b) Individual attribute components ($w_{a,k=4}$)



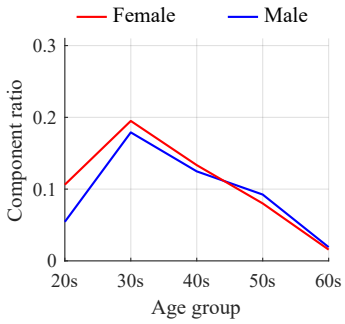
(c) Time-series components ($x_{p,k=4}$)

Figure B.11: Residence zone and attribute components of travel group $k = 4$

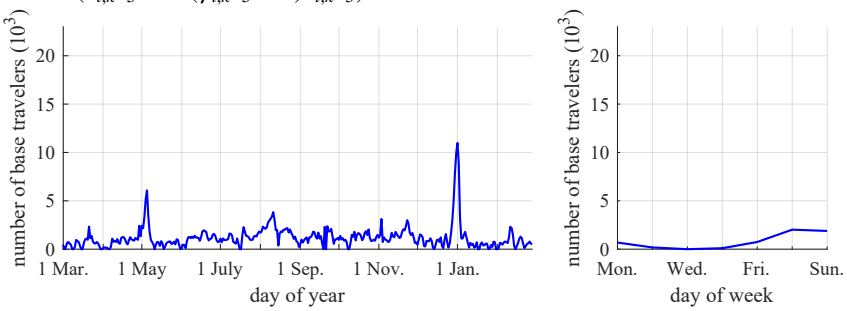
Appendix B.3. Decomposed travel group no.5



(a) Residence zone components ($v_{i,k=5}$ and $(\gamma_{i,k=5}+1)v_{i,k=5}$)



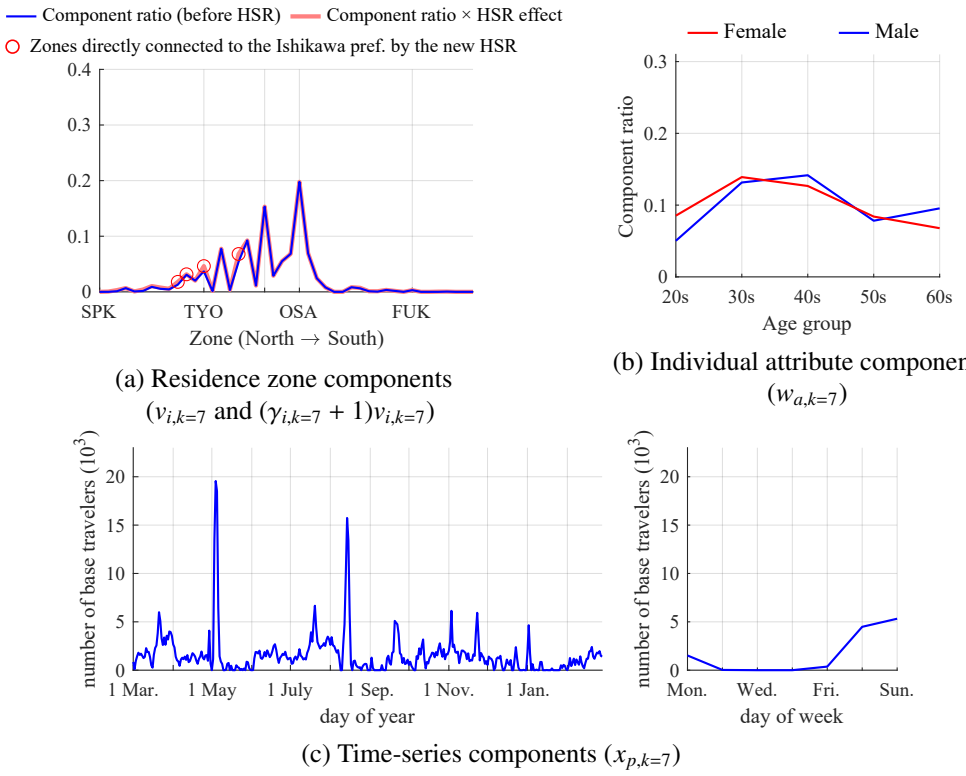
(b) Individual attribute components ($w_{a,k=5}$)



(c) Time-series components ($x_{p,k=5}$)

Figure B.12: Residence zone and attribute components of travel group $k = 5$

Appendix B.4. Decomposed travel group no.7



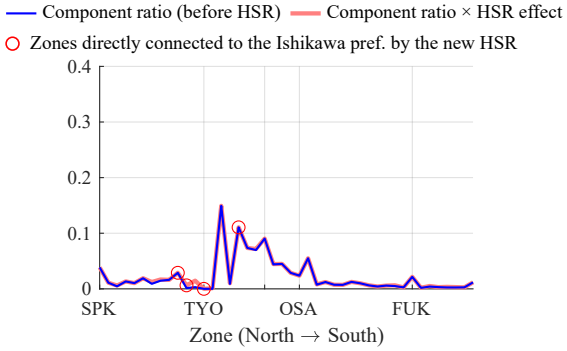
(a) Residence zone components ($v_{i,k=7}$ and $(\gamma_{i,k=7}+1)v_{i,k=7}$)

(b) Individual attribute components ($w_{a,k=7}$)

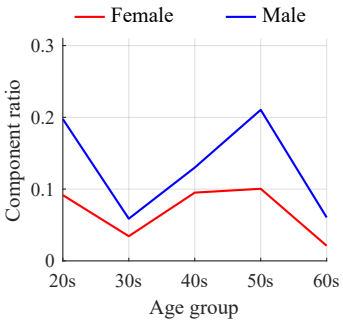
(c) Time-series components ($x_{p,k=7}$)

Figure B.13: Residence zone and attribute components of travel group $k = 7$

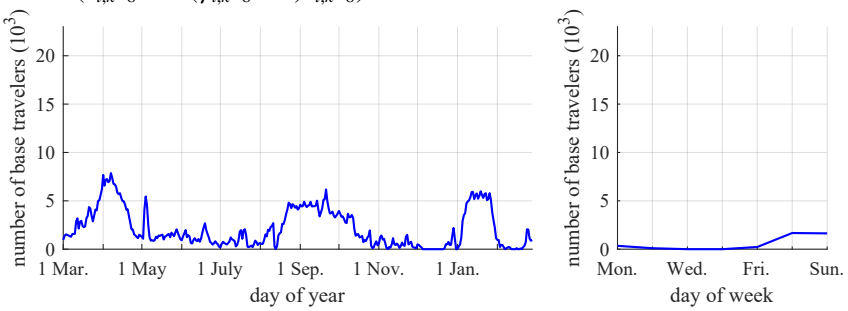
Appendix B.5. Decomposed travel group no.8



(a) Residence zone components ($v_{i,k=8}$ and $(\gamma_{i,k=8}+1)v_{i,k=8}$)



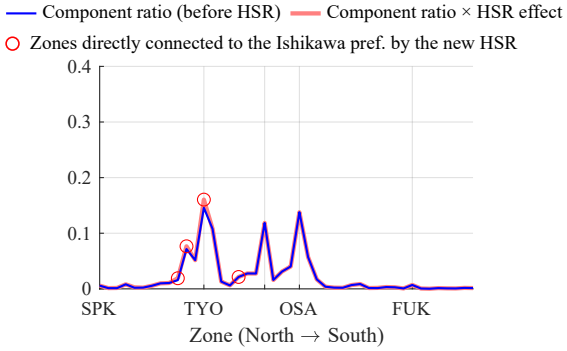
(b) Individual attribute components ($w_{a,k=8}$)



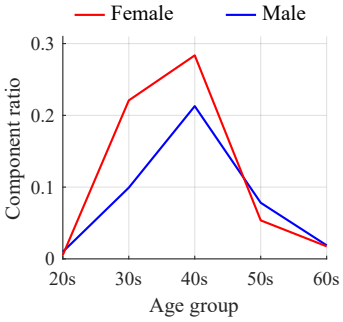
(c) Time-series components ($x_{p,k=8}$)

Figure B.14: Residence zone and attribute components of travel group $k = 8$

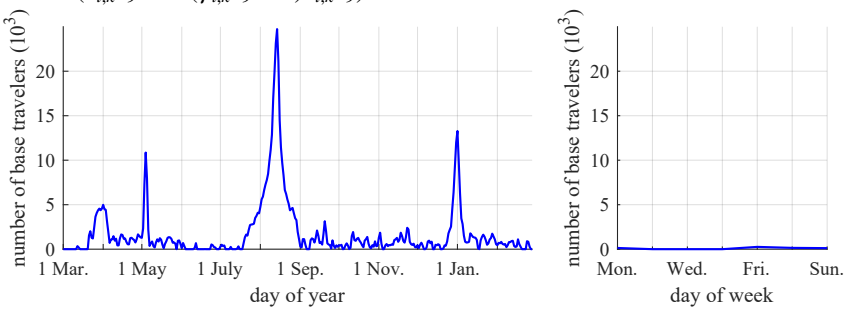
Appendix B.6. Decomposed travel group no.9



(a) Residence zone components ($v_{i,k=9}$ and $(\gamma_{i,k=9}+1)v_{i,k=9}$)



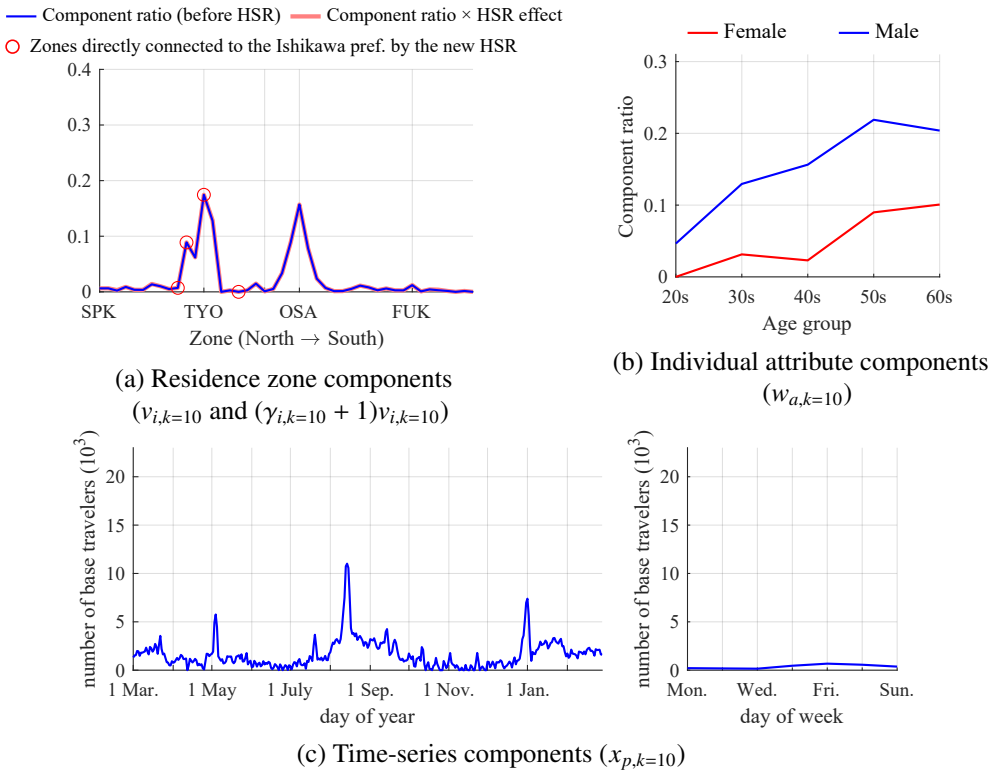
(b) Individual attribute components ($w_{a,k=9}$)



(c) Time-series components ($x_{p,k=9}$)

Figure B.15: Residence zone and attribute components of travel group $k = 9$

Appendix B.7. Decomposed travel group no.10



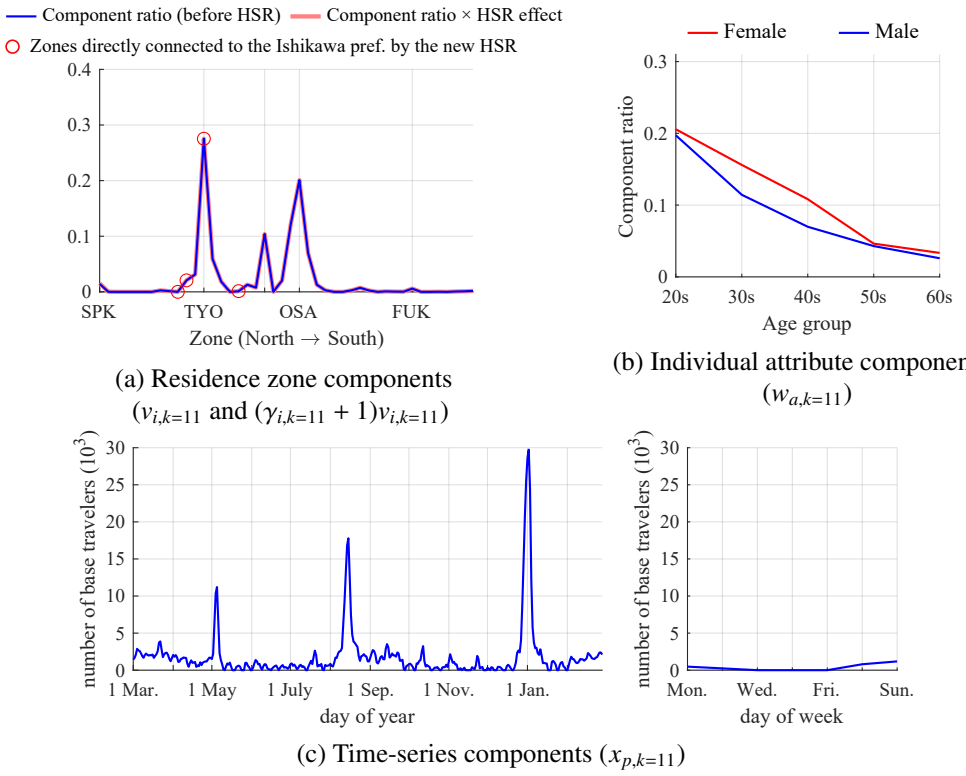
(a) Residence zone components ($v_{i,k=10}$ and $(\gamma_{i,k=10}+1)v_{i,k=10}$)

(b) Individual attribute components ($w_{a,k=10}$)

(c) Time-series components ($x_{p,k=10}$)

Figure B.16: Residence zone and attribute components of travel group $k = 10$

Appendix B.8. Decomposed travel group no.11



(a) Residence zone components ($v_{i,k=11}$ and $(\gamma_{i,k=11}+1)v_{i,k=11}$)

(b) Individual attribute components ($w_{a,k=11}$)

(c) Time-series components ($x_{p,k=11}$)

Figure B.17: Residence zone and attribute components of travel group $k = 11$